DOCUMENT RESUME

ED 115 853 CE 005 691

AUTHOR	Conroy, William G., Jr.
TITLE	Population-Sample Relationships and Data Types. Occasional Paper No. 1.
INSTITUTION	Management and Information System for Occupational Education, Winchester, Mass.
SPONS AGENCY	Massachusetts State Dept. of Education, Boston. Div. of Occupational Education.
PUB DATE	30 Jan 72
NOTE	21p.; For related documents, see ED 062 553; ED 068 646-647; ED 072 225; ED 072 228; ED 072 303-304; CE 005 687-727
EDRS PRICE	MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS	*Data; Data Bases; *Data Collection; Demonstration Projects; Information Sources; Input Output Analysis; *Management Information Systems; Sampling; State Programs; Systems Development; *Vocational Education
IDENTIFIERS	Management Information System Occupational Educa; MISOE

ABSTRACT

The paper, one of a series written as the Management Information System for Occupational Education (MISOE) was conceptualized, examines relationships between the anticipated census and sample data to be developed and maintained by the system. Two related information sources are planned: one is a census information system which includes data about every occupational education program in every school in every city and town in Massachusetts, and the other is a sample information system by occupational programs, stratified over school-type, geographical setting-type, and student characteristic-type dimensions currently operative in the State. The types and purposes of the census and sample data to be collected and maintained are described. The descriptive data (input information, process information, product information, and impact information) and analytical data (cost accounting information, process product data, cost product information, product impact information, process impact data, and cost impact data) included in the sample data system are discussed at length, defining each type and showing how each data type is related to the census. (The author suggests that Monograph Number 1, which is available as ED 062 553, be read in conjunction with this paper.) (Author/MS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED115853
CE005691

OCCASIONAL PAPER #1

POPULATION-SAMPLE RELATIONSHIPS AND DATA TYPES

William G. Conroy, Jr.

January 30, 1972

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Management and Information System for Occupational Education
1017 Main Street
Winchester, Massachusetts 01890
Telephone: 617-729-9260

TABLE OF CONTENTS

Page

Census Data 2

Sample Data 3

Descriptive Data 5

Analytical Data 13

FIGURES

1. Census Data 4

2. Product Data 7

The purpose of this brief paper is to examine relationships between the anticipated census and sample data to be developed and maintained by MISOE. None of the following is conclusion oriented but rather conceived as our thinking at this early point in developmental time. This paper will also reconsider each data type of the system in light of the sample-population relationships.

Essentially, MISOE provides a basis for describing what has occurred in occupational education in such a way that predictions for future outcomes can be optimized. Such information should provide a management tool for improving programs, allocating resources and maintaining an accountable relationship between state and local educational agencies. The very essence of the system is to provide a state-wide data base for the management of occupational education that is simultaneously useful at both state and local levels, and at the same time one which meets all the information requirements of the State Legislature, the Congress, etc. Two related information sources are planned: one is a census information system which includes data about every occupational education program in every school, in every city and town in Massachusetts, (including those programs with proprietary institutions), and the other is a sample information system by occupational programs, stratified over school-type, geographical setting-type, and student-characteristic-type dimensions currently operative in the Commonwealth. From the state's perspective, the information system will provide a careful estimate of the totality of occupational education within the state. The census data will identify specific local agencies but the sample data will not. The development of instruments which will be used to measure elements of the educational process will be administered by the state within the selected sample and made available to local educational agencies. This provides a

basis for LEAs to compare themselves to the appropriate strata within the sample and at the same time provides a basis for the state educational agency to conduct an educational audit at will.

It is important to continue to think of the system under development as not only an information system but a management system. The management and information system is designed to encourage (consistent with the Monograph #1) appropriate flexibility for decision making at the local level, while at the same time provide a comparable state-wide data base. This is difficult business. It is important that the system does not become a tool by which mindless bureaucrats can standardize educational practices in the Commonwealth.

I. CENSUS DATA

The next section of this report presents a cursory examination of the census data that will be collected and maintained by the total information system.

The purpose of census data is severalfold:

(1) To present an annual projection of enrollments, expenditures and specific performance objectives for management and accountability purposes, by program, by city or town, and by school.

(2) To present an annual (historical) description of expenditures and enrollment in occupational education for management and accountability purposes by program, by city or town, and by school.

(3) To establish a population base to draw a sample for more detailed analysis such that inferences can be made to subsets within the population.

Census data is both a planning and accountability tool. It is designed to provide the linkage between each LEA and the detailed analysis of the sample, while offering an information base for instructional management (See Figure 1).

A tricky census data requirement is to provide expenditure information for accounting and analysis, consistent with both needs. The expenditure data will be carefully analyzed in the sample and is a very real consideration for both planning and accounting. My current bias is that LEAs will have to maintain their books (census) at the program level while the sample data will deal with analysis at the objective level within programs. The tools to analyze costs at the objective level will be provided to the LEAs so that they can translate their fiscal information at the program level into a form such that they can analyze local realities in light of state information, or the State Department of Education can conduct an educational audit by selected community.

Census data does provide a base for the management of occupational education by the State Department of Education throughout a current year. Further, the census data becomes particularly useful in the light of the information provided by the sample in that it offers the basis for comparison between the sample and individual programs within schools. A rationale for all census data will be made explicit as part of the early developmental work of this project.

II. SAMPLE DATA

An appropriate sample by programs over subsets of the population about which inferences are to be made must be drawn. Represented in the

FIGURE 1

CENSUS DATA

BY JULY 1, 19 .

PLANNING DATA

ANTICIPATED ENROLLMENTS

by Program
by Level
by School Type
by City or Town
by Student Characteristics

Sex
Race
SES
I. Q.
PL 90-576 Constraints

ANTICIPATED END YEAR PERFORMANCE OBJECTIVE (1)

by Program
by Level
by School Type
by City or Town

ANTICIPATED EXPENDITURE (2)

Over-all Programs
By Program

by School
by City or Town
by PL 90-576 Constraints
by Level

BY OCTOBER 1, 19 .

REAL DATA

ENROLLMENTS

by Program
by Level
by School Type
by City or Town
by Student Characteristics

Sex
Race
SES
I. Q.
PL 90-576 Constraints

BY JUNE 30, 19 .

REAL DATA

COMPLETIONS

by Program
by Level
by School Type
by City or Town
by Student Characteristics

Sex
Race
SES
I. Q.
PL 90-576 Constraints

BY AUGUST 1, 19 .

REAL DATA

EXPENDITURES (2)

Over-all Programs
By Program

by School
by City or Town
by PL 90-576 Constraints
by Level

Assumptions:
(1) Students are enrolled in occupational programs for which they have a high probability of success.

Requirement:
(2) All expenditure data must be by source, i.e., Federal, State or local.

sample would be a variety of meaningful student-types, within each of the various occupational education programs, pursuing a fairly comprehensive battery of program objectives over appropriate levels and within representative school and city and town types. Such a sample will be identified in a way that each local community (or the State) can estimate the fit of an LEA into the stratified sample data cube. (A note to myself is that considerable work needs to be done on thinking through the process of reporting goals and objectives to allow [encourage] diversity, constrained only by knowledge.)

It is within this sample that the detailed descriptive and analytical data described in Monograph #1 will be attained. The sample will be drawn from the planning data of July (see Figure 1) and corrected by the October 1 data, to detect any major discrepencies. This will permit the gathering of information of the sample data from the very beginning of the school year. The assumption here is that everybody reading this Occasional Paper has read Monograph #1. The following section merely explores each information type described in Monograph #1 in the light of census-sample relationships.

III. DESCRIPTIVE DATA

Logically, the first data to be discussed is descriptive data. Product data will be discussed first, as it is basic to all other data and fundamental to educational management. The following is offered as a way of coding descriptive data:

D_1 - Input information

D_2 - Process information

D_3 - Product information

D_4 - Impact information

D_3 - Product Data

Product data really involves each local educational agency stipulating its product goals, that is, behavioral objectives, by program and by level. The relationship between the census and sample is that schools can compare themselves with the sample (or be compared) in terms of the specific objectives which they are offering. The other information types allow additional individualized analysis to occur; however, at this point we are dealing only with product data.

Figure 2, below, stipulates objectives in an imaginary, 3-year secondary school program. For both the census and sample, information will be available which describes the program and level, by year, at which the objectives are offered. It is assumed that these will vary by sequence offered across programs. However, it is anticipated the variance of terminal objectives within programs and across schools will be slight. Further, it is assumed that each objective offered, no matter at what level of the program, describes a competency or capability that is a part of the total educational product as it "rolls off the educational assembly line". Such a requirement will be stipulated in detail as the project develops. The sample information system keeps track of the order or sequence in which the objectives are offered, as this might, and probably will, account for some of the end program variance, but only measures product at the end of the program. That would be the twelfth grade for secondary school programs, typically; the fourteenth grade for post-secondary school programs, or at course termination for adult programs or MDTA Programs. The comparative analysis for LEAs, therefore, is at the point of program exit and it is at this point that mea-

FIGURE 2

PRODUCT DATA

SECONDARY SCHOOL

GRADE	10	11	12
	$X_{1.1}$	$X_{2.1}$	$X_{3.1}$
	$X_{1.2}$	$X_{2.2}$	$X_{3.2}$
	$X_{1.3}$	$X_{2.3}$	$X_{3.3}$

TEST AT END OF GRADE 12 FOR OBJECTIVES

$X_{1.1} - X_{3.3}$

X = OBJECTIVE

surement and analysis in terms of achievement data will occur. The information system being developed will only concern itself with measuring students in the sample at program completion and will provide the tools for measurement to the LEAs such that they can measure their own students in light of the goals they have selected. Further, the LEAs can measure, if they wish, the progress of students immediately after completion of a particular learning experience. An important p[illegible] made here is that the system being developed is not an achievement m[illegible]rocess which typifies individualized instruction projects, but a large information system into which a variety of subsystems may fit. It bears repeating that the sequence in which the objectives are offered are more a process consideration as described in Monograph #1 than product. The connectiveness between sample and census data is obviously of prime importance, as this represents a major bridge in the integrated information system.

D2 - Process Data

Given information which describes students' occupational capabilities within a sample (which is stratified over important dimensions of occupational education and is connected to each environment in which occupational education is offered) the obvious analytical question is - "What are the elements of the occupational program that account for the variance or achievement described by the product data?" For the purpose of developing process data within each sample, the following distinction is useful: (1) process elements, both structural and organizational, that are common to all objectives within a program such as organizational press, teachers' salaries, expenditure level, etc., and (2) process variables that are unique to a particular configuration of objectives; for example, the specific curriculum

elements which are related to a particular set of program objectives, like equipment, teaching strategy and instructional time (process space of the IPPI model is currently under development). Both kinds of information are necessary to account for variance, and thereby allow for prediction.

Process data should be gathered carefully during each year so that analysis might be made at the end of the particular school cycle. The instruments to describe process data will be made available to local educational agencies, so that they can determine the degree to which their programs are similar to those in the sample. This provides an important bridge or linkage between the sample and the population.

There are two interactive positions for collecting process information that will be used: (1) that which exists; (2) the research literature which indicates those elements which are most likely to have positive relationships with successful performance. The model should allow for local educational agencies to experiment with new processes in a way that information can be attained and generalized that describes the outcomes of such experimentation within the model.

D_1 - Input Data

Like all other data in the sample, the input data must provide a bridge to the general population and at the same time detailed information for careful analysis. There are two kinds of input data other than numbers of students and they have been categorized as expenditure and student characteristic data.

The student characteristic data will be considerably more complex than that which is collected as a part of the census data (see Figure 1), and the instruments for detecting these differentiations will be made

available to LEAs so that they might make comparisons to the sample data. The really important criterion in terms of determining ways of looking at individual differences is usefulness in accounting for variance. Educational findings are typically applied to groups of people rather than to individuals and, therefore, the goal is to determine which treatments are most effective for which groups of students, and it is these groups that one tries to identify. For example, it would be a goal of the information system to be able to say that the chances are seven out of ten that students within this particular range can accomplish these particular goals with that particular treatment at some specified price. If we could do just that.

It is not clear to me how expenditure information should be gathered, coded and stored. For example, is it good enough for LEAs to only keep records of cost data over all programs, and in the sample keep track of cost data by programs? Can we estimate program cost for each LEA from overall expenditure data? Should we force LEAs to (I shudder at the thought) keep books by program? How do we make sample population bridges? These problems are for the economists to work on. In the sample, however, cost data must be developed for expenditure information which costs out data not only by program but by objective. We must not only know how much money was spent within a total program, but how much was spent on a class of students to demonstrate a particular capability. This is the whole essence of product cost data to be discussed below. I am sure that one has to work with clusters of objectives here. The method must be sufficiently flexible so that its maintenance in the hands of the bureaucracy doesn't tend to standardize programs (always an eminent danger). Cost data must also be maintained in such a way that marginal as well as average costs can be determined, in addition to requirements for cost benefit analysis, to be discussed later on.

Resource allocation is obviously an important management function and cost data is of prime importance and must be easily bridgeable to the total population. May I repeat - my personal bias is to keep the cost expenditure data simple with the general population and detailed within the sample, providing tools for analysis on the part of the LEA. LEAs can then use these measuring tools to determine the relationship between their reality and the sample and these instruments might also be used by the Department of Education to conduct a periodic educational audit. I recognize this as an extremely complex area which must be responsive to analysis needs, and consistent with existing practices in the State and Feceral Government.

D_4 - Impact Data

Impact Data is simply described as the impact of occupational education on society and the student over time, which is to say follow-up information. At this point in time, I am not at all sure whether the impact data, in view of the difficulty to obtain it, should be a subsample of the sample or the total sample population, with double sampling for non-respondents. However, I suggest there should be two samples for analysis, one from which general characteristics or general information is systematically collected over time (number of jobs, job satisfaction and satisfactoriness, productivity, citizenship behavior, etc.), and a very small sample, perhaps a hundred, on whom a very careful analysis should be performed on an observational basis. There is a bunch of data you cannot get with a questionnaire that would be extremely useful to know.

Basically, the impact data is longitudinal, that is to say, it is collected on students that we know something about before they came into the program. We will know in some detail what the components of the educa-

tional program and the capabilities of the students are at program exit. With these givens, impact data allows us to; (1) explain variance, (2) make predictions, (3) modify programs, (4) allocate resources, and (5) maintain accountability. However, we are going to have to wait for this information. In the meantime, it is going to be necessary to establish something of an evolving process with impact data, such that total analysis can begin to occur instantly. This is to suggest that we visualize something of a cross-sectional treatment operating concurrently in terms of impact data, with some fairly giant assumptions of program similarity, to be phased out as longitudinal data comes along. For example, during the first year, impact data will be available in October, or shortly after graduation, or program completion, which allows careful analysis of the impact of program completors upon self and society, but there will only be limited information available about the process and cost of the program for those particular products. It will take a number of years to develop the longitudinal aspects of this system and we seem to have an obligation to provide what Mr. Kaufman calls "good enough" data along the way. Obviously, these data must be appropriately identified, and will provide useful information, but admittedly not as good as the longitudinal impact information.

The following point is emphasized. Descriptive data in the sample must be carefully gathered for analysis, and at the same time obtained in such a way that comparisons can be made by the state educational agency, and local educational agencies. The bridges between the sample and the population must be many and useful, and it should be fairly easy for LEA and SEA personnel to make comparative determinations. This is really an essential ingredient to the whole process. The next section deals with analytical data of the sample.

IV. ANALYTICAL DATA

Monograph #1 deals with six separate analytical types, although it might be that some synthesis of these data types could occur during system development. The analytical data types are listed below:

Analytical Data

A_1	COST ACCOUNTING
A_2	PROCESS PRODUCT
A_3	COST PRODUCT
A_4	PRODUCT IMPACT
A_5	PROCESS IMPACT
A_6	COST IMPACT

A_1 - Cost Accounting Information

Cost accounting information has been previously described, which is keeping track of the cost of occupational educational programs by program and program elements across school types, levels, cities and towns, etc., and by behavioral or product objectives. This is obviously both essential and difficult. We must know the cost of the programs, we must know the essence of the programs, we must know the students the programs treat, we must know the outcome of these programs, we must know the impact of students on society of these programs and all this information must be internally connected and bridgeable to the total population. All this doesn't belong here but it came to me at this moment and that way, so there it is.

A_2 - Process Product Data

This also has been pretty well described. Product data is limited to those capabilities which the student can demonstrate at the completion of the program. Process product information assumes this information is available in a reliable way and that there is a way of connecting the product data back through process data to input information. Process product information assumes that the information can be sliced by appropriate student characteristic configurations. If you will remember, the sequence of learning or of offering objectives by LEA is also available information. What process product information simply does is to describe relationships between process and product for particular student groups. It includes comparisons between student groups and between processes. It bears repeating that all information types are distinguishable by specific student characteristic types. It is my bias that these types be limited to a reasonable array. I think it is absurd to consider individualized instruction as anything that will ever happen in

American education as it just appears to be unnecessarily inefficient. I continue to subscribe to the bias that people can be individually tracked into seven or twenty classes which provides a reasonably diverse structure for prescribing educational treatments.

A_3 - Cost Product Information

Cost product information is really process product information with cost, "ground in". It simply provides a dollar estimate for process product information, obviously sliceable by student characteristic configurations. Monograph #1 describes the usefulness of this information to educational managers, and it is difficult to imagine how one could ever manage instruction without this data. Hopefully, this kind of information will detect enormous differences in the cost of preparing program completors with occupational competency, particularly in light of being able to deal with the data differentially by student types.

> The next three data types deal with impact and assume that longitudinal impact data is available. As indicated above, such information will not be available for several years, and we therefore must manufacture cross-sectional data which will be appropriately identified.

A_4 - Product Impact Information

If educational product is defined as the configuration by individual, or by classes of individuals, of occupational competencies or capabilities at the completion of the educational experience, then product impact information attempts to detect relationships between these behavioral patterns and their impact on self and society over a period of time. Since product data is

connectable to process and input data, the next three data types are sort of dominoes off this comparison.

It might be noted that we are dealing with an internal system. A separate, connectable data base that we do not have and absolutely need is information about what some describe as the validity of product data or the objectives that are being offered in occupational educational programs. This is currently arrived at by some sort of a random, not very well organized, process of advisory committees. Project CAREER (a Massachusetts development project) is attempting to get some of this information which will be coded into MISOE, but such information is not a fundamental part of the system. However, product impact data does get at this need. Over time, it will feed back information into the internal system which describes the variance of different lifestyles and occupational success patterns accounted for by the accomplishment of various product objectives. Since this data is all connectable, it can be differentiated by student input type.

A_5 - Process Impact Data

It should be obvious by this point in time that process impact data is an attempt to detect relationships. This information provides the tool to analyze the usefulness of various processes in light of impact over time.

An appropriate observation might be that product information is the essential information for the educational manager. If MISOE can reference product data or behavioral objectives by the events that cause them and their likely impact, then they become the essential management tool. Again, it is important to recognize that all of the product data is coded, not only by Office of Education codes, but cross-referenced to DOT codes.

A_6 - Cost Impact Data

Cost impact data really is something of a new data type which provides a tool for estimating relationships between cost and impact. It allows a means to analyze the benefits to society and the student of occupational education. Benefits are (by definition in Monograph #1) restricted to impact variables quantifiable by the dollar and to this must be added other kinds of information to get the total impact picture.

Also involved in the determination or estimation of this information is the need to establish a comparable sample(s) of people who have received differential, non-occupational treatment. We leave the determination of this to the economists, but indicate that this information does deal with a larger issue of whether to invest in occupational education or not, and to what extent. The objective is to arrive at a configuration such that the benefits exceed the cost in light of other investment alternatives, so that one can determine what the additional benefits of additional expenditures will be, and that these are not only known but positive, prior to the investment of the additional dollars. A separate sample must be determined by cost benefit analysis for comparison purposes at the secondary level, post-secondary level, for MDTA and adult programs and any other separate occupational education programs that enter into the analysis picture. These samples should be established at the time students begin an occupational program; for example, the ninth or tenth grade for secondary programs, the thirteenth grade for post-secondary programs, etc.

Conclusion

Fundamentally, the purpose of this paper is to re-examine data types of the information system in light of sample population relations. Other

papers will deal with developmental concerns, with Occasional Paper #2 treating developmental tasks.

This paper references Planning Chart #1, which should be concurrently considered.